

Methods of analysis for georeferenced sample counts of tarnished plant bugs in cotton

J. L. Willers · J. N. Jenkins · J. M. McKinion · Pat Gerard ·
K. B. Hood · J. R. Bassie · M. D. Cauthen

Published online: 10 October 2008
© Springer Science+Business Media, LLC 2008

Abstract The problem of analyzing georeferenced cotton pest insect samples when a large percentage of the counts are zero is examined. The use of appropriate statistical methods for their analysis is required. To demonstrate this, georeferenced samples ($n = 63$) of tarnished plant bugs (TPBs; *Lygus lineolaris* [Palisot de Beauvois] (Heteroptera: Miridae)) were analyzed by three statistical methods and the results were compared. Correlation analysis of the sample counts with 25 classes of cotton growth derived from an unsupervised classification of multispectral imagery was followed by a complete enumeration analysis comprising three scenarios. The first scenario assumed the insect samples were unstratified. A distribution of sample averages was created by complete enumeration of all combinations of samples taken four at a time. The second scenario used imagery of the cotton fields to allocate the samples among three cotton growth categories (marginal, good or best) derived by a supervised classification of the 25 unsupervised classes. The insect samples associated with these categorical habitats were completely enumerated using allocations of 4, 6, 8 or 10 samples at a time from various sample sizes to determine how different allocations affected the results. The mean was not affected, but the standard deviation decreased with increased allocation sizes in all habitats. The third scenario used the two observers and three habitat categories to create six

J. L. Willers (✉) · J. N. Jenkins · J. M. McKinion
Genetics and Precision Agriculture Research Unit, USDA - ARS, P.O. Box 5367, Mississippi State,
MS 39762, USA
e-mail: jeffrey.willers@ars.usda.gov

P. Gerard
Department of Applied Economics and Statistics, Clemson University, Clemson, SC, USA

K. B. Hood
Perthshire Farms, Gunnison, MS, USA

J. R. Bassie
Bassie Ag Service, Cleveland, MS, USA

M. D. Cauthen
Cauthen Entomological Service, Rena Lara, MS, USA

additional complete enumeration distributions by allocating four samples at a time from groups of varying sample sizes. These enumeration distributions are non-parametric estimators of the sampling distribution of: (1) the sample averages of a given sample size when samples are taken from the entire field, (2) the sample averages of a given size when samples are taken from each cotton habitat or (3) the sample averages of a given sample size from samples taken from each habitat by each observer. To support the enumeration analyses, these insect samples were analyzed further by Poisson regression models. These models showed significant differences between TPB counts by the two observers and among the habitats, whereas the observer by habitat interaction was not significant. For every combination of observer and cotton growth category, a Poisson regression model estimated the mean rate of TPB numbers. These means were similar to the corresponding modes of the complete enumeration distributions. The two non-standard methods showed that TPB numbers differed by habitat categories even though there were samples with a zero count, whereas a correlation analysis failed to identify a relationship between TPB sample counts and unsupervised habitat classes.

Keywords Sampling · Complete enumeration · Poisson regression · Remote sensing · Cotton insect management · Site-specific agriculture

Introduction

Cotton producers depend upon consultants to make many major production decisions, including insect control. Consultants traverse fields, examining plants along the way, and based upon their observations, make recommendations to treat or not to treat the crop, or to wait and examine the field again before making a decision. These decisions require considerable skill, experience and the use of efficient sampling methods (Nyrop et al. 1986; Trumble 1985; Willers et al. 2005) to assess the status of pests. Consultants and producers are highly skilled and experienced. They know what to do when insect or other agronomic problems arise, but their sampling efficiency is limited by time constraints and many fields scattered over large areas (Fig. 1). On large farms, consultants cannot spend enough time sampling any one field to obtain the large sample size required by traditional statistical theory (Karandinos 1976). As a result, pesticide applications are based on judgments involving small sample sizes that historically result in blanket applications over large groups of fields at similar times. This practice contributes to the development of pesticide resistance and increases in both environmental concerns and costs of production (Stern et al. 1959; Wilson et al. 1989).

Precision agriculture is based on the ability to measure, record and analyze georeferenced information on production activities in fields. Precision agriculture technologies can help farms to improve decisions on insect control and reduce cotton production costs (Campenella 2000). The processing of georeferenced information by Geographical Information Systems (GIS) and the use of variable-rate and yield monitor systems installed on farm equipment enable analyses of site-specific information (Boydell and McBratney 2002; Willers et al. 2008) and better management decisions (Blom et al. 2002; Midgarden et al. 1997).

Wilson (1994) reviewed procedures that provide estimates of the abundance and distribution of arthropods in cotton. Irrespective of the choice of sampling technique (Pedigo and Buntin 1994; Thompson 1992), it is difficult to describe adequately insect population sizes and their spatial distribution across a farm landscape. This task is easier if remotely

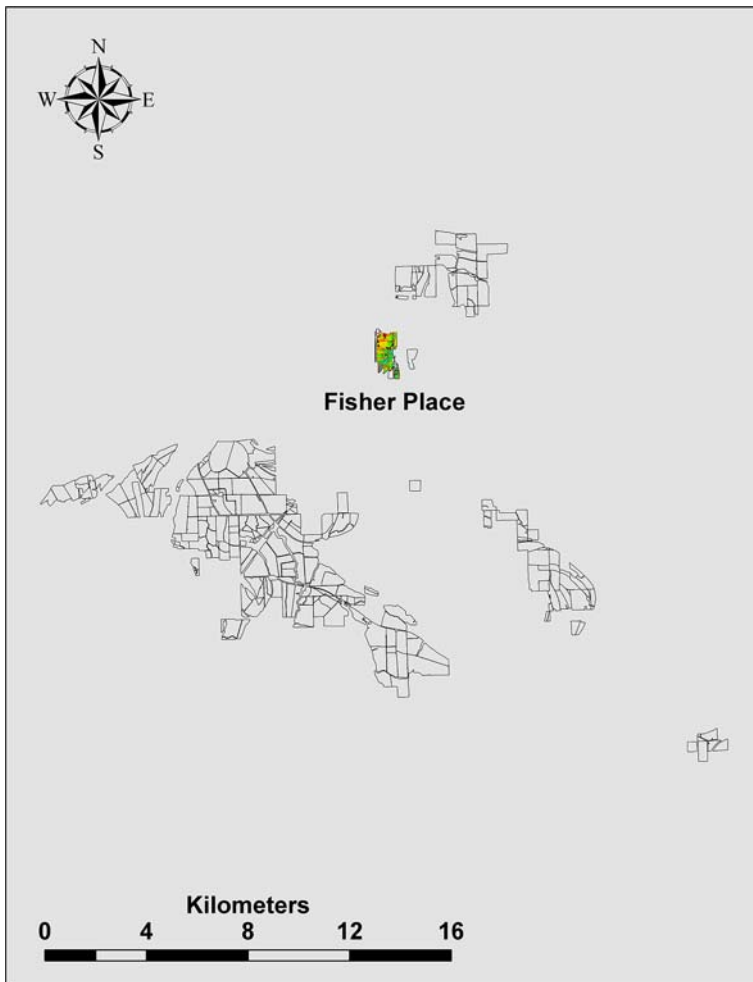


Fig. 1 Map showing location of the Fisher Place cotton fields (center) in relationship to other fields on the farm

sensed information is available to facilitate the selection of sample locations (Shaw and Willers 2006; Willers et al. 1999, 2005). However, data from georeferenced cotton insect samples require non-standard methods of analysis because the response variable (i.e. the number of insects collected in a sample) is a count (Johnston 1993; Long 1997). Methods similar to ordinary least squares (OLS), the standard regression method, can be problematic for count data (Myers and Montgomery 1997) because, according to Johnston (1993), they are not normally distributed and the variances of counts are not constant for all observations.

Complete enumeration, although not commonly used (Manly 1997), is related to resampling techniques (Confalonieri et al. 2007; Efron 1981). Some assumptions of resampling are: (1) each sample is considered an equally likely and valid representation of those that could occur with an infinitely large sample from the process being studied, (2) the sample values actually obtained are assumed to be a random sample of the total

possible and (3) it is permissible to take a random sample from the measured samples if they are independent and equally probable values through the measurement process (Iwi et al. 1999). Applications of complete enumeration in entomology are uncommon, but Willers et al. (2000, 2005) applied it to examine the effects of sample unit size and mixed population densities on sample averages.

Poisson regression represents an application of generalized linear models (Long 1997; Piegorsch and Bailer 2005). Such models are a refinement of traditional linear models in that the mean of a population depends on a linear predictor through a non-linear link function, and the probability distribution of the response variable is a member of the exponential family of distributions (Johnston 1993; Nelder and Wedderburn 1972). A few entomological applications only of count models are currently available in the literature (e.g. Candy 2000; Desouhant et al. 1998; Hall 2000), but none involves site-specific cotton pest management.

As data analysis of insect samples by standard statistical methods is problematic, our objective was to use one standard and two non-standard methods to test the hypothesis: Tarnished plant bug (TPB; *Lygus lineolaris* [Palisot de Beauvois]) numbers increase as the quality of the cotton habitats improves. Using a typical data set from a commercial field, we used correlation as the standard method and complete enumeration and Poisson regression as the non-standard methods and compared the results.

Materials and methods

This paper integrates applications and concepts from several scientific disciplines. To assist the reader, an Appendix is included with a glossary of terms.

Classification of imagery

A group of nine cotton fields, called the Fisher Place (Fig. 1, Center), located on a large farming operation were selected for this experiment. The multispectral imagery (Jensen 2000) was obtained by InTime, Inc. (Cleveland, MS, <http://www.gointime.com>) on 10 July 2004. The fields and the 3 band (near-infrared, red and green) imagery were provided by Perthshire Farms, Gunnison, MS. The spatial resolution of the imagery was a 2×2 m pixel registered to Universal Transverse Mercator (UTM) coordinates for Zone 15N, using the WGS84 datum (Bugayevskiy and Snyder 1995).

In previous applications of multispectral imagery for developing site-specific recommendations for TPB management (Willers et al. 1999, 2005), the georeferenced insect samples were collected before or at the start of peak bloom. The scouting map used to select sample locations in these earlier studies was derived from the normalized difference vegetation index (NDVI) (Elvidge and Chen 1995; Rouse et al. 1974). However, we have learned that NDVI is ineffective for cotton insect scouting during and after peak bloom. This study took place when the best cotton habitat category was near the end of peak bloom. We overcame the ineffectiveness of NDVI at this stage of growth by using the unsupervised classifier of ERDAS® Imagine® 8.7 (Norcross, GA) to create the field scouting map. The unsupervised classifier used all 3 spectral bands of the image to assign each pixel to one of 26 classes (Pouncey et al. 1999; Richards and Jia 1999). Class 0 represented the landscape outside the field boundaries, whereas a number between 1 and 25 represented the unique unsupervised class value assigned to a pixel within the field boundaries. With 25 classes, each assigned a colour within five main colors (red, orange,

yellow, green and blue), a detailed map of the variation in cotton growth was derived. This level of detail was necessary for the field observers to discern subtle differences in cotton growth at various geographic locations. The observers used this map to select their sample locations and to derive the three habitat categories used later for the non-standard statistical analyses.

Collection of field samples

Detailed descriptions and assumptions of the sampling plan in this study are in Willers et al. (2005, 2006). An important feature of geographical sampling is that the sample unit size is considered to be a small area with several adjacent cotton plants on which to count TPB, rather than the traditional sample unit sizes of single plants, numbers of sweeps, or drop cloth samples (Musser et al. 2007; Wilson 1994). This emphasis on geographic areas of fields is in distinct contrast to the traditional focus of insect sampling that compares an estimate of the field average to a threshold value without considering the location of samples. A georeferenced sampling scheme to record the number of cotton insects should be designed to enable the number of insects per unit area to be compared within and among different regions of the cotton field. Both observers used large sample unit sizes to minimize the chance that any given sample would have a zero count; this is likely where a particular habitat category has a low TPB density (Willers et al. 2005). Otherwise, the TPB counts obtained for the three methods of analysis are typical of samples obtained by consultants from commercial cotton fields.

Two observers collected 63 georeferenced samples (50 by sweep-net and 13 by drop cloth) of TPB counts from the nine commercial cotton fields on the 19th and 22nd of July 2004 (Fig. 2). The farm consultant (Observer 1) sampled the fields on both dates, whereas the research entomologist (Observer 2) sampled them only on the second date. These observers comprised the two levels of the explanatory variable named OBSERVER in the non-standard statistical analyses. Using hand-held GPS receivers, both observers determined the coordinate locations of their sample sites to a horizontal accuracy of 3–5 m. The coordinates of the sample locations were not post-processed (Kennedy 1996) to improve their horizontal accuracies.

GIS processing of samples and imagery

A 10 m buffer was used with the GIS to determine the zonal average and maximum (Theobald 2003) of the unsupervised class values of the pixels (Pouncey et al. 1999) and these values were retained as new attributes for each TPB sample location. For each sample, the final attribute table (Theobald 2003) contained these zonal statistics together with information such as date and time of sample, sample identifier, its geographic coordinates, observer and total number of TPB by adding together the numbers of adults and nymphs.

Methods of statistical analysis

An iterative supervised classification procedure (Pouncey et al. 1999; Richards and Jia 1999) re-grouped the 25 unsupervised classes into three ordinal categories for use with the complete enumeration and Poisson regression analyses. These three categories correspond to different levels of cotton vigor and plant density and comprised the field habitats for TPB. Other fields, at other times, may be represented better by different numbers of habitat

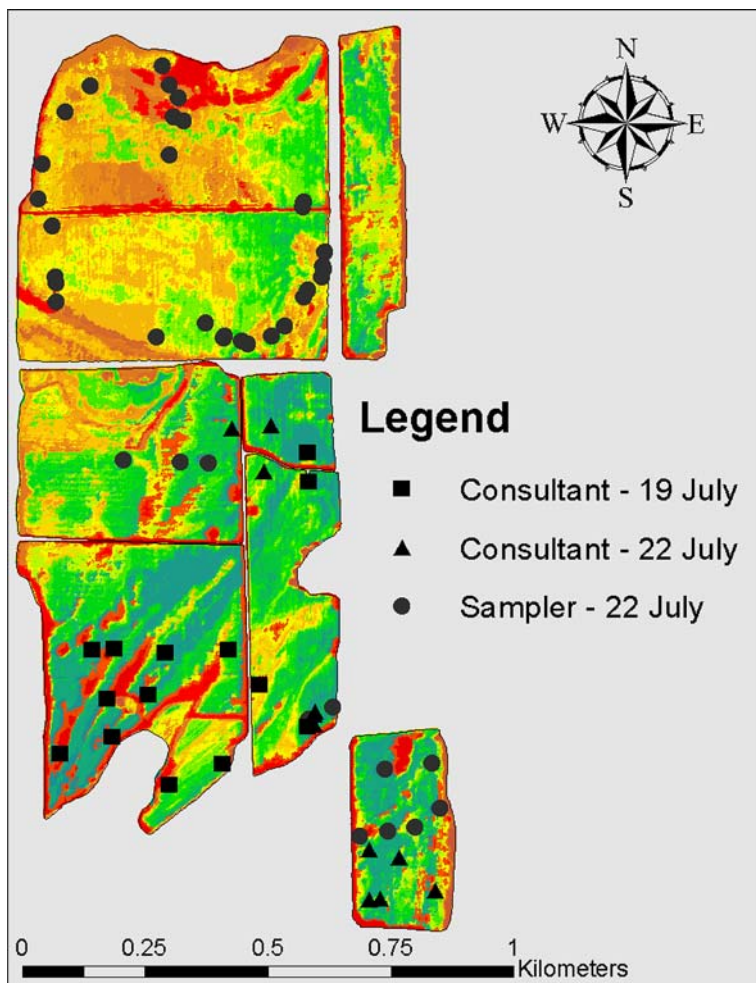


Fig. 2 Map showing locations of the sample sites by date for two observers for the Fisher Place cotton fields

categories (Willers et al. 2005). The unsupervised classification class value limits for the marginal habitat category were ≥ 1 and < 9 and ≥ 20 . Unsupervised classes ≥ 20 were labeled as marginal habitat because they represented cotton plant populations of mixed low, medium and high vigor, with low density and of varying row lengths where there were no plants (i.e. skips). The cotton from areas in classes < 9 differs from that in classes ≥ 20 only insofar as the plants are uniformly of low vigor and are not mixed with medium or high vigor plants. The unsupervised classification class value limits for the good habitat were ≥ 9 and < 17 and for the best habitat they were ≥ 17 and < 20 . The marginal, good and best categories comprised the three levels of the HABITAT explanatory variable used with the non-standard analyses.

Three scenarios were established for statistical analysis. The first scenario was without stratification, the second comprised samples stratified by HABITAT (three levels) and the third scenario had samples stratified by HABITAT and OBSERVER (six combinations).

Table 1 The listing of sampling vectors: (a) pooled over observers and arranged into one group assuming that no ancillary information from imagery is available, (b) stratified into three groups according to categories of cotton growth established from image information and (c) stratified into six groups by category and observer, using the geographic coordinates of the samples. The sample size of each vector is provided in the last column

Label	Sampling vector	<i>n</i>		
(a)				
Unstratified	0 0 1 0 1 0 0 0 1 0 0 1 0 1 0 2 0 1 2 1 2 1 0 0 2 0 3 0 1 0 1 2 2 0 2 4 5 1 4 0 1 0 1 1 0 0 2 0 1 1 2 0 1 1 0 4 0 0 1 3 2 0 0	63		
(b)				
Best	0 2 4 5 1 4 0 2 0 1 1 2 0 1 1 0 4 0 0 1 3 2 0 0	24		
Good	0 0 2 0 1 2 1 2 1 0 0 2 0 3 0 1 0 1 2 2 0 1 1 0	24		
Marginal	0 0 1 0 1 0 0 0 1 0 0 1 0 1 1	15		
Label	Sampling vector	<i>n</i>		
	Observer 1	<i>n</i>	Observer 2	<i>n</i>
(c)				
Best	0 2 0 1 1 2 0 1 1 0 4 0 0 1 3 2 0 0	18	0 2 4 5 1 4	6
Good	0 1 1 0	4	0 0 2 0 1 2 1 2 1 0 0 2 0 3 0 1 0 1 2 2	20
Marginal	0 1 1	3	0 0 1 0 1 0 0 0 1 0 0 1	12

Normality and correlation analyses

The TPB sample counts for the unstratified and selected stratified scenarios were tested for normality with the SAS[®] Univariate procedure (SAS[®] Institute, Ver. 9.1), with a significance level of $\alpha=0.05$. The Pearson product moment correlation coefficient was determined between the TPB counts and the zonal average and zonal maximum of the unsupervised class values of the pixels contained in the buffer of each sample (SAS[®] Proc Corr; SAS[®] Institute, Ver. 9.1). Scatter plots of these results were produced with SAS/Graph[®] Proc Gplot.

Complete enumeration analyses

A complete enumeration (Manly 1997, p. 8) procedure was used to analyze the TPB sample counts for each scenario. Complete enumeration distributions were created using appropriate selections of the various sampling vectors with the following combinatorial rule (Beyer 1968):

$$\binom{n}{m} = C(n, m) = \frac{n!}{m!(n-m)!}. \quad (1)$$

For a specified sampling vector (Table 1a, b or c), this rule determined the number of combinations (C) of n samples taken m at a time. Unlike other resampling algorithms (Efron 1981; Manly 1997) that randomly ‘resample the samples’, this approach obtains the frequency distribution of the estimated means of all $C(n, m)$ unique combinations without randomization. The result represents a non-parametric estimate of the simple random sampling distribution of TPB sample averages of size m for the different scenarios. Using the programming language C, the $C(n, m)$ combinations were ordered lexicographically

and the mean TPB count for each combination was determined (Willers et al. 2000). Most of the source code was obtained from a website maintained by Sloane (2007). Histograms were created with SAS[®] Proc Capability to summarize results from the three scenarios.

The first scenario used allocations of size $m = 4$ where $n = 63$. When the OBSERVER and HABITAT explanatory variables (Thompson 1992; Willers et al. 2005) are not taken into account, the average of each combination represents an estimate of the ‘field average’ of TPB. For the second scenario, several choices for m were selected (4, 6, 8 and 10) for a sensitivity analysis of m using sampling vectors of various sizes (n) pooled over observers, but stratified by habitat. The mean, standard deviation, minimum and maximum statistics were computed for the 13 complete enumeration distributions of the first and second scenarios. For the 6 complete enumeration distributions of the third scenario, the allocation size was fixed at $m = 4$, whereas the values of n (the observed sample sizes) varied with the different combinations of habitat and observer. Summary statistics for the third scenario were determined by inspecting its histograms.

Poisson regression analyses

Insect samples result in count data. For analyses of counts, Poisson regression approaches are improvements over OLS methods (Long 1997; Seavy et al. 2005; Slymen et al. 2006). However, if the counts are over-dispersed (Long 1997), a Poisson regression model analysis will under-estimate the standard errors and over-estimate the test statistics (Allison 1999; Slymen et al. 2006). Over-dispersion (Agresti 1996) occurs when the observed variance is larger than the nominal variance for a specified statistical distribution. Over-dispersion was tested using the full Poisson regression model (Johnston 1993), where all the levels of OBSERVER, HABITAT and their interactions were specified. If the counts are over-dispersed, another kind of count model must be used (Long 1997; Myers and Montgomery 1997; Ridout et al. 1998).

Based on Slymen et al. (2006), the Poisson regression model is described as follows. Let y_i represent the TPB count ($y_i, i = 0, 1, 2, \dots$) for the i th sample. Let \mathbf{x}_i represent a vector of covariates named HABITAT and/or OBSERVER and let β be a vector of regression coefficients to be estimated. The Poisson model (Allison 1999; Long 1997; Stokes et al. 2000) is given as:

$$\text{Prob}(y_i|\mathbf{x}_i) = \exp(-\lambda_i) \lambda_i^{y_i} / y_i!, \quad (2)$$

where

$$\lambda_i = \exp(\mathbf{x}_i' \beta). \quad (3)$$

The expected value of y_i given \mathbf{x}_i is λ_i (Slymen et al. 2006) and the variance (var) of the observations equals the mean (Allison 1999; Johnston 1993; Long 1997) or:

$$E(y) = \text{var}(\lambda) = \lambda. \quad (4)$$

The strategy for fitting a Poisson model to these TPB sample counts was similar to the approach described by Johnston (1993). First, the full model was fitted to the sample counts as:

$$\lambda_{jk} = \exp(b_0 + \text{OBSERVER}_j + \text{HABITAT}_k + \text{OBSERVER} * \text{HABITAT}_{jk}), \quad (5)$$

where b_0 is the intercept, $j = 1, 2$ and $k = 1, 2, 3$. As a first step, the model given by Eq. 5 was used to determine whether the interaction term is significant. For georeferenced TPB

counts, if this interaction is significant, then one or both effects represent sources of random sampling error and have no useful interpretation as main effects by this model (Eq. 5). Reasons for the significant interaction should then be identified and alternative methods of analysis considered.

On the other hand, if the interaction is non-significant, the next step (Johnston 1993) is to fit a reduced model without any explanatory variables:

$$\lambda = \exp(b_0), \quad (6)$$

followed by addition of the explanatory variable OBSERVER:

$$\lambda_j = \exp(b_0 + \text{OBSERVER}_j), \quad (6a)$$

and then HABITAT to obtain:

$$\lambda_{jk} = \exp(b_0 + \text{OBSERVER}_j + \text{HABITAT}_k). \quad (6b)$$

From this set of candidate models, the last step is to select the model that provides the best fit with the smallest number of terms and use it for further analyses. The fit of the model at each step is monitored by comparing changes in deviance (Johnston 1993). The goodness of fit of the selected model is tested by the likelihood ratio (LR) test (Schabenberger and Pierce 2002):

$$\text{LR} = 2(\ln L_{\text{Unrestricted}} - \ln L_{\text{Restricted}}) \sim \chi^2(J), \quad (7)$$

where J is the difference in the number of regressors between the unrestricted and restricted models (Park 2005).

Results

Traditional statistical analyses

Tests of normality

The Shapiro–Wilk statistic, W , was significant ($W = 0.782$, $p < 0.0001$) for the unstratified TPB samples. Tests for normality of the samples by habitat categories (pooling over OBSERVER) were significant for the marginal ($W = 0.630$, $p < 0.0001$), good ($W = 0.827$, $p = 0.0008$) and best ($W = 0.831$, $p = 0.0010$) cotton growth categories. Similarly, tests for normality of the sample counts (pooling over HABITAT) were significant for both OBSERVER 1 and 2 ($W = 0.786$, $p = 0.0001$; $W = 0.786$, $p = 0.0001$, respectively). The strength of significance of these results indicates that these sample counts are not normally distributed and methods based on assumptions of normality would be inappropriate (Allison 1999; Long 1997). To emphasize this fact, the correlations between the sample counts and the zonal average and the zonal maximum of the sample sites are discussed next.

Correlation of TPB samples with unsupervised habitat classes

Using the unsupervised class values, Fig. 3 shows the scatter plot of total TPB numbers against the zonal average of the pixels contained within the buffer at the corresponding sample location. The estimated correlation coefficient ($r = 0.21$) between these TPB

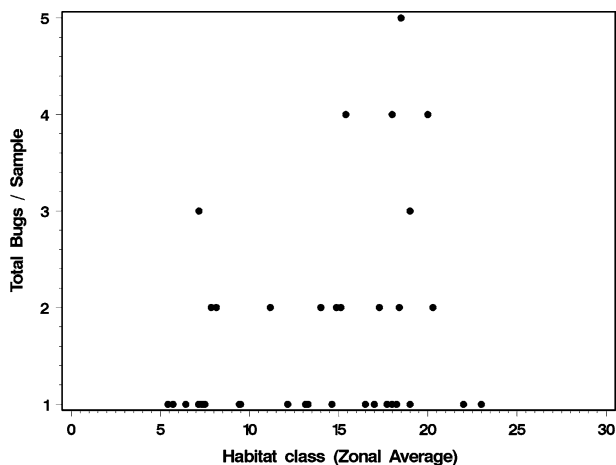


Fig. 3 Scatter plot of tarnished plant bug sample counts versus the zonal average of the habitat classes of pixels within a 10 m buffer centered on the sample site location

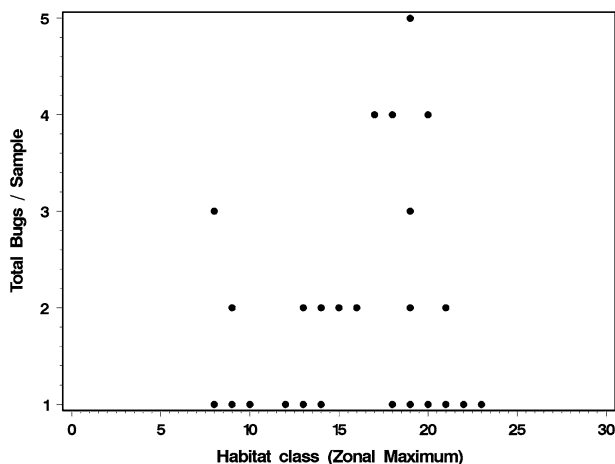


Fig. 4 Scatter plot of tarnished plant bug sample counts versus the zonal maximum of the habitat class pixels within a 10 m buffer centered on the sample site location

counts and the zonal average is not significant ($p = 0.09$). The zonal maximum of the unsupervised class values of pixels at the sample locations and the TPB counts (Fig. 4) also shows no linear trend ($r = 0.11$, $p = 0.37$). Even if these correlations were significant, they are so small that few commercial consultants would prepare a site-specific prescription for TPB control with these results.

The lack of a relationship using a standard method of analysis is to be expected because these TPB counts were not normally distributed. Also, the unsupervised classes do not represent an ordered list of cotton growth vigor. These results mean that non-standard methods of analysis are required to obtain evidence for geographical differences in the abundance of TPBs.

Complete enumeration

Over a third of the pooled TPB sample counts (Table 1a) are zeros. However, using complete enumeration to estimate the distribution of sample averages of size m minimizes the impact of samples whose count was zero. For instance, the complete enumeration distribution for the unstratified scenario (Fig. 5) represents the sampling distribution of the field average for sample averages of size four by simple random sampling (Thompson 1992; Willers et al. 2005). This distribution describes the expected results of sample averages for the third week of July 2004, as if an observer used a simple random sampling design without geographic stratification and repeatedly collected only four samples. In the long term, 3% of the results would average zero TPB, 27% would average 0.5 TPB, 33% would average 1 TPB and so forth. From this distribution, the chance that a large sample average (i.e. ≥ 3 TPB) will be observed is small if a habitat map is unavailable to guide the selection of sample sites. Other summary statistics for this scenario are given in Table 2, row 1.

The allocation of samples (Table 1b) among the three habitat categories (marginal, good and best) for sample sizes of $m = 4$, with pooling across observers, results in different complete enumeration distributions (Fig. 6). The histograms show that the averages tend to increase with improvement in cotton habitat quality. This trend suggests that TPB numbers are not uniform across these fields and that there is a potential for yield loss in the best cotton habitat regions of these fields. The result for the unstratified scenario, which makes no geographical distinction among the samples, shows little evidence of any potential problem among these fields at this time. The histogram of the unstratified example is provided for comparison (bottom panel, Fig. 6).

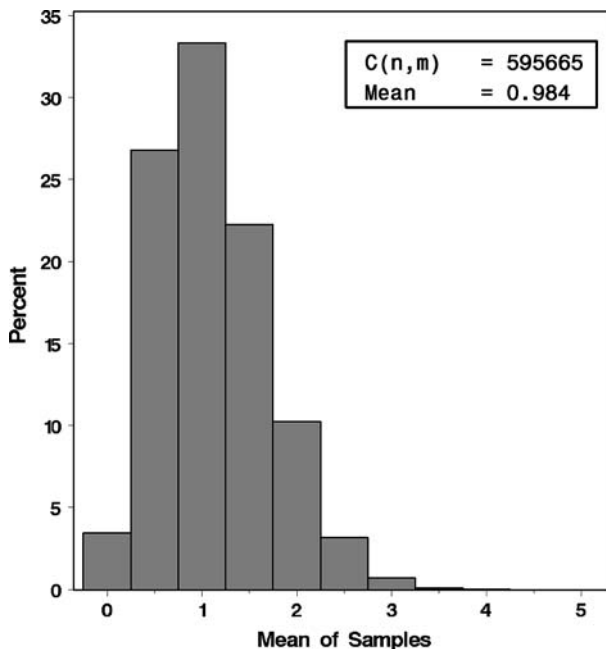


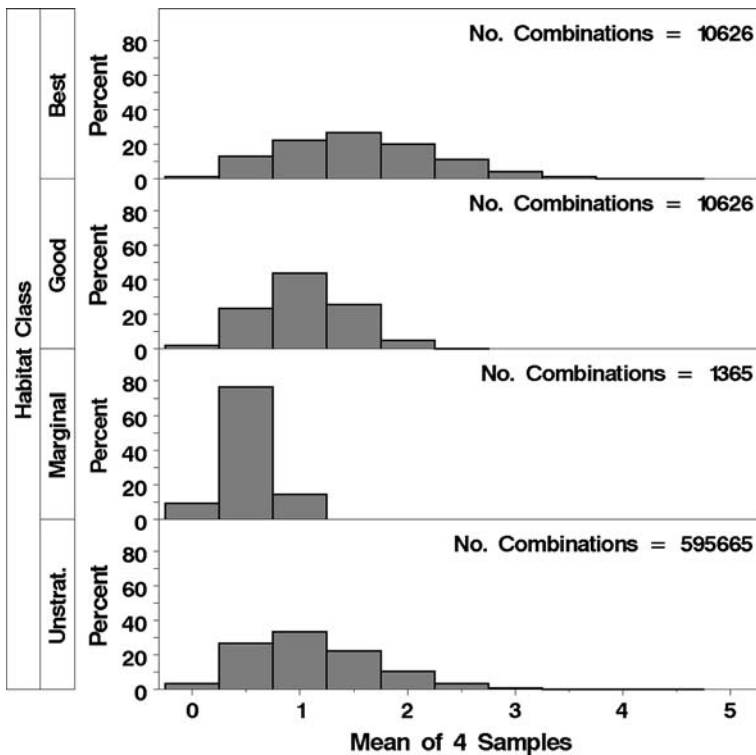
Fig. 5 Complete enumeration distribution of $n = 63$ unstratified tarnished plant bug sample counts. With an allocation size of $m = 4$, the number of sample average combinations ($C(63,4)$) is indicated along with the overall mean

Table 2 Summary statistics from the complete enumeration distributions of means for N combinations of four groupings of sampling vectors (i.e. unstratified, best, good and marginal) presented in Table 1a and b

Label	N^a	Mean	Standard deviation	Minimum ^b	Maximum ^b
Unstratified ($m = 4$)	595,665	0.98	0.58	0	4.25
Best ($m = 4$)	10,626	1.42	0.71	0	4.25
Best ($m = 6$)	134,596	1.42	0.55	0	3.67
Best ($m = 8$)	735,471	1.42	0.45	0	3.25
Best ($m = 10$)	1,961,256	1.42	0.38	0.10	2.90
Good ($m = 4$)	10,626	0.92	0.42	0	2.25
Good ($m = 6$)	134,596	0.92	0.33	0	2.18
Good ($m = 8$)	735,471	0.92	0.27	0	2.00
Good ($m = 10$)	1,961,256	0.92	0.22	0	1.80
Marginal ($m = 4$)	1,365	0.40	0.22	0	1.00
Marginal ($m = 6$)	5,005	0.40	0.16	0	1.00
Marginal ($m = 8$)	6,435	0.40	0.12	0	0.75
Marginal ($m = 10$)	3,003	0.40	0.09	0.10	0.60

^a The value for number of combinations (N) is determined by $C(n, m)$ where n = the number of samples in the four sampling vectors and m is set to 4, 6, 8 or 10

^b These minimum and maximum values are an average according to the argument m in $C(n, m)$

**Fig. 6** Complete enumeration distributions of sample averages for the pooled observers across three habitat categories in comparison to the unstratified sampling distribution, with an allocation size of $m = 4$

A sensitivity analysis examined effects on the sampling distribution of averages among the habitats pooled over observers for other allocation sizes of $m = 6, 8$ and 10 . Within habitat categories, this analysis (Table 2, Rows 2–13) shows little difference among these means compared to allocations of size four, whereas the estimated standard deviations decrease. Differences in the standard deviation among the habitat groupings for identical choices of m (Table 2) support Johnston's (1993) assertion that for counts, the variances are not constant for all observations.

Based on the sensitivity analysis for different sizes of m (Table 2), we chose allocations of size four only to compute the enumeration distributions (Figs. 7 and 8) for the sampling vectors of the third scenario (Table 1c). Both observers' distributions have identical modes and maxima for the marginal habitat, but differ for both the good and best habitats. The greatest difference is for the good cotton habitat category. The unstratified and stratified distributions of each observer result in different numbers of combinations because of different sample sizes in the three habitats. For the two better habitats, the distribution of the averages of sample allocations for Observer 2 tends to have larger ranges and modes than for Observer 1. Reasons for these differences are discussed further below. Unlike the unstratified case, comparisons among these stratified distributions indicate that large sample averages (i.e. ≥ 3 TPB) are more prevalent in regions with the best cotton.

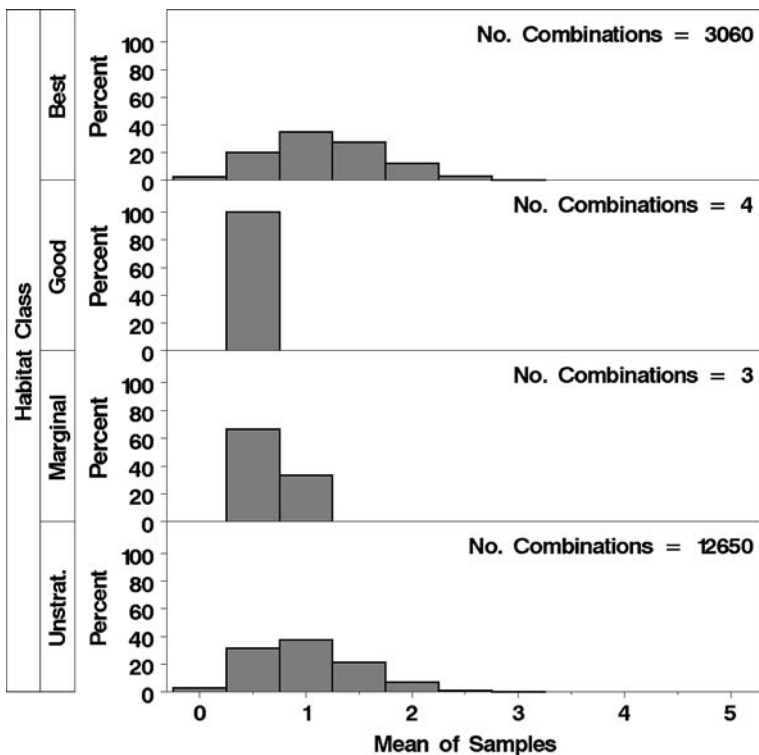


Fig. 7 Complete enumeration distributions of sample averages for the farm consultant (OBSERVER 1) for three habitat categories in comparison to his unstratified sampling distribution, with an allocation size of $m = 4$

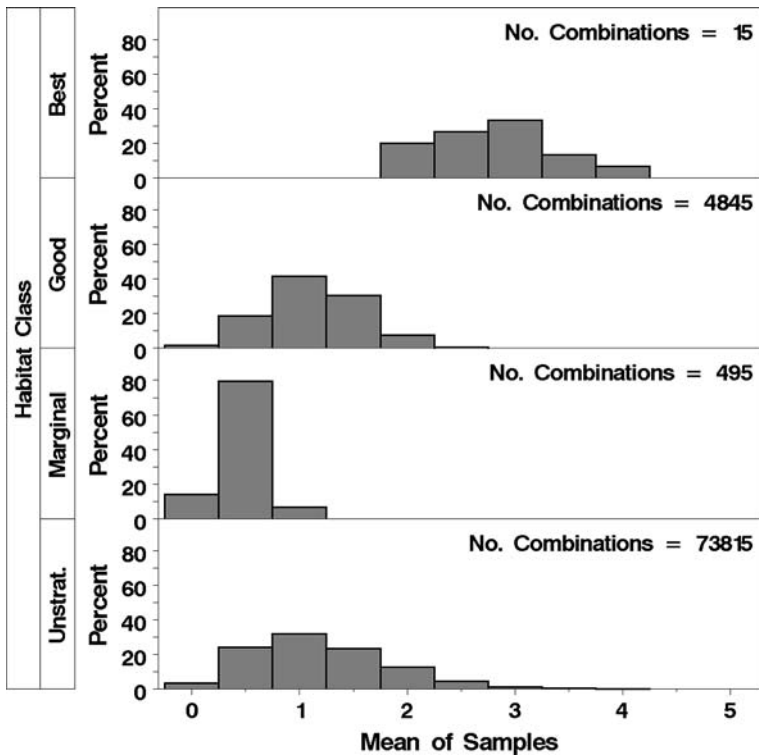


Fig. 8 Complete enumeration distributions of sample averages for the researcher (OBSERVER 2) for three habitat categories in comparison to his unstratified sampling distribution, with an allocation size of $m = 4$

Table 3 Analysis of variance table for the full Poisson regression model

Source	Likelihood ratio statistics for type 3 analysis ^a					
	Num. df ^b	Den. df ^c	F-value	Pr > F	χ^2	Pr > χ^2
OBSERVER	1	57	0.68	0.412	0.68	0.408
HABITAT	2	57	5.57	0.006	11.14	0.004
OBSERVER * HABITAT	2	57	1.27	0.289	2.54	0.281

^a Refer to the software documentation for Proc Genmod in SAS[®] for further discussion of a type 3 analysis

^b Numerator degrees of freedom

^c Denominator degrees of freedom

Poisson regression analyses

Table 3 gives the type 3 (an option for SAS[®] Proc Genmod) analysis of variance table for the full model (Eq. 5) and shows no significant interaction between OBSERVER and HABITAT. The interpretations of this result for this collection of TPB counts are that there is no variability due to the individuals collecting the sample (Musser et al. 2007) and that the habitat categories were specified properly by the supervised classification process.

With Poisson regression, the test to determine if the data are over-dispersed is accomplished by fitting Eq. 5 with the PSCALE option (Piegorsch and Bailer 2005) in SAS[®]. For these TPB counts, the estimated scale parameter is 1.03 (Poisson Model, Eq. 1, Table 4).

Table 4 Parameter estimates and assessment criteria for goodness of fit for various Poisson regression models

Model description	Parameter estimates	Standard error	$p > \chi^2$	Goodness of fit assessment criteria			
				Log-likelihood	Deviance	Pearson χ^2	df
<i>Poisson model Eq. 5</i>				−50.0	72.1	60.2	57
INTERCEPT	0.981	0.257	0.000				1
OBSERVER 1	−0.981	0.353	0.005				1
OBSERVER 2	0.000	0.000	−				0
HABITAT 1	−2.079	0.575	0.000				1
HABITAT 2	−0.981	0.345	0.004				1
HABITAT 3	0	0	−				0
OBSERVER 1 * HABITAT 1	1.674	0.958	0.080				1
OBSERVER 1 * HABITAT 2	0.288	0.840	0.732				1
OBSERVER 1 * HABITAT 3	0	0	−				0
OBSERVER 2 * HABITAT1	0	0	−				0
OBSERVER 2 * HABITAT 2	0	0	−				0
OBSERVER 2 * HABITAT 3	0	0	−				0
SCALE	1.03	0					
<i>Poisson model Eq. 6</i>				−63.0	92.3	90.4	62
INTERCEPT	−0.016	0.127	0.900				1
SCALE	1.000	0	−				
<i>Poisson model Eq. 6a</i>				−43.0	91.8	89.1	61
INTERCEPT	0.051	0.191	0.788				1
OBSERVER 1	−0.179	0.321	0.576				1
OBSERVER 2	0	0	−				0
SCALE	1.21	0					
<i>Poisson model Eq. 6b</i>				−50.6	74.7	63.2	59
INTERCEPT	0.878	0.255	0.001				1
OBSERVER 1	−0.794	0.319	0.013				1
OBSERVER 2	0	0	−				0
HABITAT 1	−1.678	0.483	0.000				1
HABITAT 2	−0.869	0.313	0.005				1
HABITAT 3	0	0	−				0
SCALE	1.03	0					
<i>Poisson model Eq. 8</i>				−79.1	−	−	−
INTERCEPT ^a	−3.248	0.883	0.000				−
OBSERVER ^a	0.787	0.299	0.008				−
HABITAT ^a	0.848	0.216	0.000				−

^a Estimated by Proc Countreg rather than Proc Genmod in SAS[®] because Poisson Model Eq. 8 required a different model matrix (the \mathbf{x}_i in Eq. 3) than Poisson Models (5–6b) to allow for exponentiation of its parameters (Johnston 1993; Long 1997). See text

If the TPB samples were from a Poisson process, the scale parameter would be 1.00 (Agresti 1996; Stokes et al. 2000). As the estimated scale parameter is close to unity and Eq. 4 is satisfied, we conclude that an analysis by Poisson regression is appropriate.

To determine the simplest, best fitting Poisson model (Johnston 1993), the deviance statistic of the fitted full model (Poisson Model, Eq. 5, Table 4) is compared to the deviance statistics of several other models. The deviance of the reduced model (Poisson Model, Eq. 6, Table 4) decreases slightly from 92.3 to 91.8 when the OBSERVER effect is added (Poisson Model, Eq. 6a, Table 4). It decreases considerably (from 92.3 to 74.8) when both OBSERVER and HABITAT are included (Poisson Model, Eq. 6b, Table 4). The greater changes in deviance with the addition of HABITAT indicate that its effect is more important than OBSERVER in explaining these data. The goodness of fit of Poisson Model, Eq. 6b, (Table 4) is assessed by Eq. 7 with $J = 3$ degrees of freedom. Since the result is strongly significant ($p < 0.0001$), we conclude that this model provides an acceptable fit to these georeferenced TPB counts.

Since Poisson Model Eq. 6b (Table 4) has the smallest deviance without interaction terms, it is selected to determine the expected probabilities (Long 1997) of observing 0, 1, 2, 3, 4 or 5 TPBs in a sample for the six OBSERVER by HABITAT combinations. These probabilities (Table 5) indicate that the chance of obtaining a zero count decreases while larger counts become more probable with increasing quality in cotton habitat. These trends can be summarized by an alternative formulation of Eq. 6b which is:

$$\lambda = \exp(b_0 + \text{OBSERVER} + \text{HABITAT}). \quad (8)$$

With this model (Poisson Model Eq. 8, Table 4), the mean increase in TPB numbers between OBSERVERs and across the HABITAT categories can be determined (Johnston 1993; Long 1997). For example, with the effects of HABITAT kept constant, OBSERVER 2 generally found 2.2 TPB for every 1 TPB found by OBSERVER 1. This is found by exponentiation of the parameter estimate for OBSERVER (i.e. $\exp(0.787)$). As for HABITAT, the TPB counts increase by 2.3 TPB (i.e. $\exp(0.848)$) with each increase in habitat level from marginal to best, while keeping the effect of OBSERVER constant.

The Poisson regression analyses verify the complete enumeration results because the estimates of Poisson means (Table 5) for the six observer by habitat combinations are close to the modal values of the comparable complete enumeration distributions (Figs. 7 and 8).

Table 5 Estimated tarnished plant bug means and probabilities of occurrences for count values in samples from two observers over three cotton growth categories predicted by Poisson regression Model, Eq. 6b of Table 4

Effect ^a	Predicted TPB mean ($\hat{\lambda}$)	Tarnished plant bug count per sample (probability ^b)					
		0	1	2	3	4	5
OBSERVER 1, Marginal	0.203	0.816	0.166	0.017	0.001	0.000	0.000
OBSERVER 1, Good	0.456	0.634	0.289	0.066	0.010	0.001	0.000
OBSERVER 1, Best	1.087	0.337	0.367	0.199	0.072	0.020	0.004
OBSERVER 2, Marginal	0.449	0.638	0.287	0.064	0.010	0.001	0.000
OBSERVER 2, Good	1.009	0.365	0.368	0.185	0.062	0.016	0.003
OBSERVER 2, Best	2.405	0.090	0.217	0.261	0.209	0.126	0.060

^a See Johnston 1993; Seavy et al. 2005

^b Probabilities may not sum to one due to rounding error or because probabilities estimated for count events >5 are not presented

Both methods provide effective ways of analyzing georeferenced samples from commercial cotton fields in spite of the occurrence of zero counts, differences in sample location and different sample sizes among the habitat categories. Results from the two non-standard methods provide a way forward to develop better site-specific pesticide prescriptions (e.g. Dupont et al. 2000) for the management of TPB.

Discussion

If there is no difference in the counts of insects from samples grouped by habitat, the complete enumeration distributions will have similar centers and shapes. Similarly, the habitat effect of a Poisson regression model will be non-significant. If the insect counts from cotton habitats differ, the shape and or location of the centers of complete enumeration distributions will shift and the habitat effect of a Poisson regression model will be significant. The histograms of enumerated averages and the Poisson regression model estimates of probabilities of different count values provide the analytical capability to examine how habitat categories (or other explanatory variables relevant to the study) affect the shape of the distribution. The importance of considering differences in shapes of sample counts by habitat categories was first noted when Bayesian methods were used to analyze insect counts (Willers and Akins 2000). That work and the results of this study suggest that interpreting the shapes of enumeration distributions (or Poisson probability distributions) derived from TPB sample counts facilitates geographical or site-specific decisions for pest management. These results also prepare the way for testing hypotheses about site-specific action thresholds (Fleischer et al. 1999; Stern et al. 1959) for different habitat categories and how they should differ over the production season.

The two observers in this study demonstrate differences in approaches to sampling while determining the insect infestation levels in an area. Although both observers collected their samples by a similar process, their sampling pattern and intensity was different. Nevertheless, the enumeration distributions for the two observers can be considered as valid representations of TPB numbers in the habitat categories of these cotton fields during the study in 2004. In spite of the notable difference in sample size and site selection by the two observers (Fig. 2), the difference in TPB averages between observers was not a reflection of their sampling skills (i.e. each has more than 30 years of experience in sampling cotton). This difference in TPB averages was due to the location of their sample sites in each habitat category, which reflected the endogenous variation in TPB abundance within the habitats. The map of habitat categories (Fig. 2) shows that the consultant (Observer 1) sampled toward the southern area where the percentage of the best cotton habitat is greater. The map also shows that this observer, unlike Observer 2, generally selected sample sites in the lower range of the best category (e.g. class 17, Fig. 9).

Neither the consultant nor the researcher spent much time where the TPB population was expected to be the largest (Willers et al. 1999, 2005), i.e. in the best cotton (deep blue hues on the map of Fig. 9). Observer 1 was familiar with recent spatio-temporal trends in TPB numbers over these fields and focused his sampling efforts toward the lower ranges of the best category in an effort to find the boundary class breakpoint between it and the good habitat category. His previous sampling experiences indicated that TPB populations were greater in the upper set of the unsupervised classes representing the best habitat and he needed no further samples to confirm this. When Observer 2 chose sites within the upper range of the best cotton (e.g. unsupervised class 20; Fig. 9) and discovered that 3 out of 6 samples had counts ≥ 3 , he concluded that TPB numbers were already large and it was

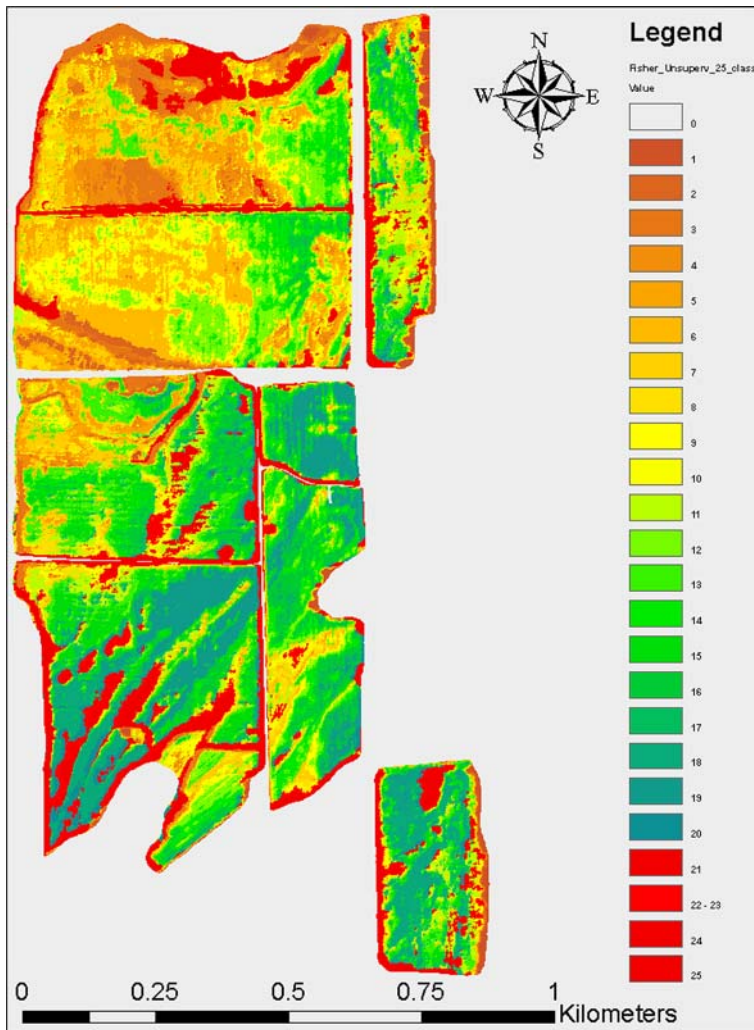


Fig. 9 Illustration of the geographical distribution of the 25 unsupervised classes derived from the three bands of the multispectral image for the Fisher Place cotton fields. This figure illustrates the scouting map used by the observers to select their sample sites

unnecessary to collect more samples (Willers et al. 2005). Observer 2 focused his efforts across the marginal categories to obtain a large enough sample size to demonstrate that larger sample sizes in that category were also unnecessary. This result was reflected by Observer 1, who chose only 3 sample locations in the marginal habitats (Table 1c) and also found very small numbers of TPBs. The sample size and locations used by Observer 1 in the marginal habitat category indicates that Observer 2 spent too much time sampling there.

The sample size and site selection of Observer 2 in the upper unsupervised classes of the good habitat was similar to the sampling rationale used by the field consultant for his sample site selection in the lower unsupervised classes of the best habitat. In other words,

based upon the unsupervised classification map, both observers actually spent their sampling time learning where to establish a breakpoint between the good and best habitat categories. This behavior by both observers suggests that additional research is necessary to determine time-saving methods that better establish class limits among habitats. Overall, the selection of 25 sample sites by Observer 1 was as informative in describing the TPB risk geographically (Willers et al. 2005) in these cotton fields as the combined sample size of 63.

Some investigators have criticized results of this type of study because the sample sites were not distributed evenly across the unsupervised habitat classes (or the supervised habitat categories). Consequently, they maintain that the complete enumeration distributions for the unstratified and stratified scenarios favor one habitat more than another. In a commercial field environment, requiring consultants to have equal sample sizes in all habitat categories is unrealistic because it does not recognize the constraints on their sampling time and it undervalues their judgment. Our results show that both complete enumeration and Poisson regression provide excellent summaries of geographical sample counts in spite of the fact that sample sizes between the observers and among the habitat categories were not equivalent. This finding indicates that equal allocation of sample sites among habitat categories is not required for commercial field applications.

The unsupervised classification map described the geographical differences in crop vigor, which enabled the two observers to select their sample locations independently. The re-grouping of that map into supervised classification habitat categories facilitated the analyses of the sample counts by non-standard methods. The analyses of geographical sample counts are valid for only a short time and need to be done frequently because other circumstances such as weather and previous pesticide spray patterns influence insect abundance and dispersion (Fleisher et al. 1999; Stern et al. 1959). The temporal nature of insect sampling requires producers and consultants to have a quick and accurate sampling method to make management decisions. It is feasible to meet this requirement if habitat classification maps derived from remote sensing imagery are available (Willers et al. 1999, 2005), but it is also necessary to analyze sample counts rapidly by geographical regions. On commercial farms, complete enumeration is a more appropriate approach than the fitting of a Poisson regression model which requires artistic involvement by an experienced analyst. Therefore, our next step is to develop farm friendly applications of complete enumeration methods.

Conclusions

Statistics derived from complete enumeration distributions or Poisson regression models of georeferenced TPB samples are better for use in TPB pest management than those obtained by standard methods of analysis (OLS). Results from Poisson regression analyses confirmed the validity of complete enumeration analyses. Both methods modelled the sample counts as functions of categorical explanatory variables. Poisson regression is a more appropriate tool for the research or industrial investigator who has the time to collect larger sample sizes, whereas the graphical display of complete enumeration distributions is preferred for farm applications where time is limited and sample sizes are smaller. The non-standard methods showed that, although the two observers allocated their sampling efforts differently among the habitat categories, the analyses of both sets of counts produced similar results about the geographic abundance of TPB. In summary, the hypothesis that TPB abundance differs among cotton habitats was substantiated by both the complete

enumeration and Poisson regression analyses, whereas the correlation analysis failed to show any relationship. This contrast in results between a standard and two non-standard analysis methods indicates that for questions involving georeferenced sampling information the choice of analysis technique is a major consideration. This finding has important implications in precision agriculture for the site-specific management of pests on a valuable crop such as cotton.

Acknowledgements Appreciation is expressed to Ms. Debbie Boykin and Dr. Gordon Snodgrass (ARS, USDA, Stoneville, MS), Mr. Ronald E. Britton (ARS, USDA, Mississippi State), Drs. Wes Burger and Bronson Strickland (Department of Wildlife and Fisheries, Mississippi State University), Dr. Fred Musser (Department of Plant Pathology and Entomology, Mississippi State University), Dr. Matt Kramer (ARS, USDA, Beltsville, MD), Dr. Jiaying Wu (Department of Plant and Soil Sciences, Mississippi State University) and Yongfeng Zhao (Department of Mathematics and Statistics, Mississippi State University) for their helpful discussions or reviews during the preparation of this manuscript. We thank the editor and two anonymous reviewers for other helpful suggestions. Financial support was provided by Advanced Spatial Technologies for Agriculture (ASTA-322-298) and the USDA Area-Wide Tarnished Plant Bug Management Project (thru ARS CRIS Project 6406-21610-006-00D). Approved for publication as Journal Article No. J-11245 of the Mississippi Agricultural and Forestry Experiment Station, Mississippi State University.

Disclaimer Mention of a trademark, proprietary product, or vendor does not constitute guarantee or warranty of the product by the U.S. Department of Agriculture and does not imply its approval to the exclusion of other products or vendors that may also be suitable.

Appendix

To develop the analysis processes and concepts of this paper, the following definitions are given:

- (1) *Buffer*: A GIS polygon, typically a circle of a specified radius used to define a region within which specialized spatial processing operations are restricted.
- (2) *Complete enumeration*: A resampling method where all possible combinations of a sampling vector are derived by a combinatorial rule expressed as:

$$C(n, m) = n! / m!(n - m)!$$

which defines the number of combinations (Beyer 1968) of n distinct items taken m at a time. The combinatorial rule was applied to enumerate all allocations (m) of a chosen sample size of (n) tarnished plant bug (TPB) samples without repeating the selection of sample counts contained in any prior combination.

- (3) *Complete enumeration distribution*: The histogram that describes the frequency of events (i.e. the mean of samples taken m at a time) that result from the complete enumeration of the sample vector (a set of n samples).
- (4) *Count data*: Integer value for the number of items of interest counted in a sample of a specified sample unit size.
- (5) *Habitat category*: A collection of pixels whose attribute values belong to the same category and represent cotton whose phenology of growth is more similar than other categories (Willers et al. 2005). In a specialized usage, it can be a synonym for the GIS concept of a zone in raster processing (Theobald 2003).
- (6) *Over-dispersion*: This occurs when count data show greater variability than is predicted by the assumed mean-variance relationship (Hinde and Demétrio 1988;

- Long 1997). If ignored, the selected statistical model may underestimate the standard errors and result in inappropriate inferences on the estimated regression parameters.
- (7) *Peak bloom*: That period during the production season when the plants produce the maximum number of open flowers per plant.
 - (8) *Poisson regression*: A regression technique based upon the Poisson probability mass function where the estimates of parameters are derived by maximum likelihood methods.
 - (9) *Raster format*: A GIS file format where the features are represented as a collection of cells of the same size and shape for a specified number of rows and columns. Each cell is georeferenced and only one integer valued attribute is assigned to each cell (Theobald 2003).
 - (10) *Sample*: A set of observations obtained with a particular sampling method (Willers et al. 2005) at a geographic location in a farm field. The response, or dependent variable, is the number (or count) of insects found in the sample.
 - (11) *Sample size*: The number of samples (n) collected by an observer while using a specified sampling method.
 - (12) *Sample unit size*: In this study, the area of ground represented by a single sample obtained with a specified sampling method (Willers et al. 2000).
 - (13) *Sample vector*: The collection of n samples collected by an observer during a specific period of time. The sampling vector can be unstratified (no ancillary data associated with the sample) or geographically stratified (ancillary data associated with the sample). Important ancillary data (Willers et al. 2005) associated with a sample are the sampler, its geographic coordinate (or label), and the cotton habitat class from which the sample was obtained. This concept must not be confused with vector data in GIS processing.
 - (14) *Simple random sample*: The simplest sampling design in which a sample site is chosen by a randomization process from the set, or collection, of all possible sites and where each site has an equally probable chance of being selected.
 - (15) *Stratified, simple random sample*: A sample design derivative where the samples are stratified according to some type of classification of remote sensing information that determines the number and geographical extent of the different strata. A simple random sampling design in each stratum determines the summary statistics of each stratum (Willers et al. 2005). This differs from a stratified sampling design where summary statistics for a population of interest are derived using samples from all the strata.
 - (16) *Supervised classification*: A classification of image pixels closely controlled by the analyst (Pouncey et al. 1999). In this process, the analyst selects pixels that represent patterns or land cover features that are recognized, or identified, with help from other sources, such as aerial photos, ground truth data, or maps. Knowledge of the data, and of the classes desired, is required before classification.
 - (17) *Unsupervised classification*: A computer-automated process that depends on the data itself for the definition of classes (Pouncey et al. 1999). It enables the analyst to specify some parameters that the computer uses to uncover statistical patterns inherent in the data.
 - (18) *Vector format*: A GIS file format where geographical features are represented as points and that uses their x - y coordinates to construct points, lines and areas (Theobald 2003).

- (19) *Zonal average of a buffer*: The average value of the attribute of interest for all pixels whose centers lie within the boundary of the buffer polygon (Theobald 2003).
- (20) *Zonal maximum of a buffer*: The maximum value of the attribute of interest from the set of pixels whose centers lie within the boundary of the buffer polygon (Theobald 2003).

References

- Agresti, A. (1996). *An introduction to categorical data analysis*. New York: John Wiley & Sons.
- Allison, P. D. (1999). *Logistic regression using the SAS system: Theory and application*. Cary: SAS[®] Institute Inc.
- Beyer, W. H. (1968). *Handbook of tables for probability and statistics* (2nd ed.). Cleveland: CRC Press.
- Blom, P. E., Fleischer, S. J., & Smilowitz, Z. (2002). Spatial and temporal dynamics of Colorado potato beetle (Coleoptera: Chrysomelidae) in fields with perimeter and spatially targeted insecticides. *Environmental Entomology*, 31, 149–159.
- Boydell, B., & McBratney, A. B. (2002). Identifying potential within-field management zones from cotton-yield estimates. *Precision Agriculture*, 3, 9–23. doi:10.1023/A:1013318002609.
- Bugayevskiy, L. M., & Snyder, J. P. (1995). *Map projections. A reference manual*. Philadelphia: Taylor and Francis.
- Campanella, R. (2000). Testing components toward a remote-sensing-based decision support system for cotton production. *Photogrammetric Engineering and Remote Sensing*, 66, 1219–1227.
- Candy, S. G. (2000). The application of generalized linear mixed models to multi-level sampling for insect population monitoring. *Environmental and Ecological Statistics*, 7, 217–238. doi:10.1023/A:1009662915320.
- Confalonieri, R., Acutis, M., Bellocchi, G., & Genovese, G. (2007). Resampling-based software for estimating optimal sample size. *Environment Model Software*, 22, 1796–1800. doi:10.1016/j.envsoft.2007.02.006.
- Desouhant, E., Debouzie, D., & Menu, F. (1998). Oviposition pattern of phytophagous insects: On the importance of host population heterogeneity. *Oecologia*, 114, 382–388. doi:10.1007/s004420050461.
- Dupont, J. K., Campanella, R., Seal, M. R., Willers, J. L., & Hood, K. B. (2000). Spatially variable insecticide applications through remote sensing. In P. Duggar & D. Richter (Eds.), *2000 Proceedings of the Beltwide Cotton Conferences*, vol. 2 (pp. 426–429). Memphis, TN: National Cotton Council.
- Efron, B. (1981). Nonparametric estimates of standard error: The jackknife, the bootstrap and other methods. *Biometrika*, 68, 589–599. doi:10.1093/biomet/68.3.589.
- Elvidge, C. D., & Chen, Z. (1995). Comparison of broad-band and narrow-band red and near-infrared vegetation indices. *Remote Sensing Environment*, 54, 38–48. doi:10.1016/0034-4257(95)00132-K.
- Fleischer, S. J., Blom, P. E., & Weisz, R. (1999). Sampling in precision IPM: When the objective is a map. *Phytopathology*, 89, 1112–1118. doi:10.1094/PHYTO.1999.89.11.1112.
- Hall, D. B. (2000). Zero-inflated Poisson and binomial regression with random effects: A case study. *Biometrics*, 56, 1030–1039. doi:10.1111/j.0006-341X.2000.01030.x.
- Hinde, J., & Demétrio, C. G. B. (1988). Overdispersion: Models and estimation. *Computational Statistics and Data Analysis*, 27, 151–170. doi:10.1016/S0167-9473(98)00007-3.
- Iwi, G., Millard, R. K., Palmer, A. M., Preece, A. W., & Saunders, M. (1999). Bootstrap resampling: A powerful method of assessing confidence intervals for doses from experimental data. *Physics in Medicine and Biology*, 44, N55–N62.
- Jensen, J. R. (2000). *Remote sensing of the environment: An earth resource perspective*. Upper Saddle River: Prentice-Hall.
- Johnston, G. (1993). SAS[®] software to fit the General Linear Model. In *Proceedings of the Thirteenth Annual SAS Users Group International Conference*. Cary, NC: SAS[®] Institute Inc. http://support.sas.com/rnd/app/papers/papers_da.html. Accessed 2 September 2008.
- Karandinos, M. G. (1976). Optimum sample size and comments on some published formulae. *Bulletin of the Entomological Society of America*, 22, 417–421.
- Kennedy, M. (1996). *The global positioning system: An introduction*. Chelsea: Ann Arbor Press, Inc.
- Long, J. S. (1997). *Regression models for categorical and limited dependent variables*. Thousand Oaks: Sage Publications.
- Manly, B. F. J. (1997). *Randomization, bootstrap, and Monte Carlo methods in biology* (2nd ed.). London: Chapman and Hall.

- Midgarden, D., Fleischer, S. J., Weisz, R., & Smilowitz, Z. (1997). Site-specific integrated pest management impact on development of esfenvalerate resistance in Colorado Potato Beetle (Coleoptera: Chrysomelidae) and on densities of natural enemies. *Journal of Economic Entomology*, 90, 855–867.
- Musser, F., Stewart, S., Bagwell, R., Lorenz, G., Catchot, A., Burris, E., et al. (2007). Comparison of direct and indirect sampling methods for tarnished plant bug (Hemiptera: Miridae) in flowering cotton. *Journal of Economic Entomology*, 100, 1916–1923. doi:[10.1603/0022-0493\(2007\)100\[1916:CODAIS\]2.0.CO;2](https://doi.org/10.1603/0022-0493(2007)100[1916:CODAIS]2.0.CO;2).
- Myers, R. H., & Montgomery, D. C. (1997). A tutorial on generalized linear models. *Journal of Quality Technology*, 29, 274–291.
- Nelder, J. A., & Wedderburn, R. W. M. (1972). Generalized linear models. *Journal of the Royal Statistical Society. Series A*, 135, 370–384. doi:[10.2307/2344614](https://doi.org/10.2307/2344614).
- Nyrop, J. P., Foster, R. E., & Onstad, D. W. (1986). Value of sample information in pest control decision making. *Journal of Economic Entomology*, 79, 1421–1429.
- Park, H. M. (2005). Regression models for event count data using SAS, STATA, and LIMDEP. <http://www.indiana.edu/~statmath/stat/all/count/count.html>. Accessed 2 September 2008.
- Pedigo, L. P., & Buntin, G. D. (Eds.). (1994). *Handbook of sampling methods for arthropods in agriculture*. Boca Raton: CRC Press.
- Piegorsch, W. W., & Bailer, A. J. (2005). *Analyzing environmental data*. Chichester: Wiley and Sons.
- Pouncey, R., Swanson, K., & Hart, K. (Eds.). (1999). *ERDAS field guide* (5th ed.). Atlanta: ERDAS.
- Richards, J. A., & Jia, X. (1999). *Remote sensing digital image analysis. An introduction* (3rd ed.). Berlin: Springer-Verlag.
- Ridout, M., Demétrio, C. G. B., & Hinde, J. P. (1998). Models for count data with many zeros. In *Proceedings of the XIXth International Biometric Conference* (pp. 179–192). Cape Town: International Biometrics Society.
- Rouse Jr. J. W., Haas, R. H., Deering, D. W., Schell, J. A., & Harlan, J. C. (1974). *Monitoring the vernal advancement and retrogradation (greenwave effect) of natural vegetation*. NASA/GSFC Type III Final Rpt. Greenbelt, MD: Goddard Space Flight Center.
- Schabenberger, O., & Pierce, F. J. (2002). *Contemporary statistical models for the plant and soil sciences*. Boca Raton: CRC Press.
- Seavy, N. E., Quader, S., Alexander, J. D., & Ralph, C. J. (2005). Generalized linear models and point count data: Statistical considerations for the design and analysis of monitoring studies. In C. J. Ralph & T. D. Rich (Eds.), *Bird conservation implementation and integration in the Americas* (pp. 744–753). Albany: Pacific Southwest Research Station, Forest Service, U.S. Department of Agriculture. USDA Forest Service Gen. Tech. Rep. PSW-GTR-191.
- Shaw, D. R., & Willers, J. L. (2006). Improving pest management with remote sensing. *Outlooks Pest Management*, 17, 197–201. doi:[10.1564/17oct03](https://doi.org/10.1564/17oct03).
- Sloane, N. J. A. (2007). The on-line encyclopedia of integer sequences. <http://www.research.att.com/~njas/sequences/>. Accessed 2 September 2008.
- Slymen, D. J., Ayala, G. X., Arredondo, E. M., & Elder J. P. (2006). A demonstration of modeling count data with an application to physical activity. *Epidemiologic Perspectives and Innovations*, 3:3. <http://www.epi-perspectives.com/content/3/1/3>. Accessed September 2, 2008. doi:[10.1186/1742-5573-3-3](https://doi.org/10.1186/1742-5573-3-3).
- Stern, V. M., Smith, R. F., van den Bosch, R., & Hagen, K. S. (1959). The integrated control concept. *Hilgardia*, 29, 81–101.
- Stokes, M. E., Davis, C. S., & Kock, G. G. (2000). *Categorical data analysis using the SAS® system* (2nd ed.). Cary: SAS® Institute Inc.
- Theobald, D. M. (2003). *GIS concepts and ARCGIS methods*. Fort Collins: Conservation Planning Technologies.
- Thompson, S. K. (1992). *Sampling*. New York: Wiley-Interscience.
- Trumble, J. T. (1985). Implications of changes in arthropod distribution following chemical applications. *Researches on Population Ecology*, 27, 277–285. doi:[10.1007/BF02515466](https://doi.org/10.1007/BF02515466).
- Willers, J. L., & Akins, D. C. (2000). Sampling for tarnished plant bugs in cotton. *Southwest Entomologist*, Suppl no. 23, 39–57.
- Willers, J. L., Ladner, W. L., McKinion, J. M., & Cooke, W. H. (2000). Application of computer intensive methods to evaluate the performance of a sampling design for use in cotton insect pest management. In G. A. Milliken (Ed.), *Proceedings of the 2000 Kansas State University Conference on Applied Statistics in Agriculture* (pp. 119–133). Manhattan, KS: Kansas State University.
- Willers, J. L., Jenkins, J. N., Ladner, W. L., Gerard, P. D., Boykin, D. L., Hood, K. B., et al. (2005). Site-specific approaches to cotton insect control. Sampling and remote sensing analysis techniques. *Precision Agriculture*, 6, 431–452. doi:[10.1007/s11119-005-3680-x](https://doi.org/10.1007/s11119-005-3680-x).

- Willers, J. L., McKinion, J. M., & Jenkins, J. N. (2006). Remote sensing, sampling and simulation applications in analyses of insect dispersion and abundance in cotton. In C. Aguirre-Bravo, P. J. Pellicane, D. P. Burns, & S. Draggan (Eds.), *Monitoring Science and Technology Symposium: Unifying Knowledge for Sustainability in the Western Hemisphere* (pp. 879–886). USDA Forest Service Proceedings RMRS-P-42CD. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. http://www.fs.fed.us/rm/pubs/rmrs_p042.pdf. Accessed 2 September 2008.
- Willers, J. L., Milliken, G. A., Jenkins, J. N., O'Hara, C. G., Gerard, P. D., Reynolds, D. B., et al. (2008). Defining the experimental unit for the design and analysis of site-specific experiments in commercial cotton fields. *Agricultural Systems*, 96, 237–249. doi:10.1016/j.agsy.2007.09.003.
- Willers, J. L., Seal, M. R., & Luttrell, R. G. (1999). Remote sensing, line-intercept sampling for tarnished plant bugs (Heteroptera: Miridae) in Mid-South cotton. *Journal of Cotton Science*, 3, 160–170.
- Wilson, L. T. (1994). Estimating abundance, impact, and interactions among arthropods in cotton agroecosystems. In L. P. Pedigo & G. D. Buntin (Eds.), *Handbook of sampling methods for arthropods in agriculture* (pp. 475–514). Boca Raton: CRC Press.
- Wilson, L. T., Sterling, W. L., Rummel, D. R., & De Vay, J. E. (1989). Quantitative sampling principles in cotton IPM. In R. E. Frisbie, K. M. El-Zik, & L. T. Wilson (Eds.), *Integrated pest management systems and cotton production* (pp. 85–119). New York: John Wiley & Sons.